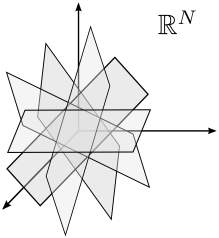
Cosparse Low-dimensional Signal Modelling

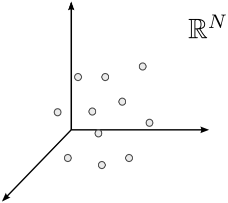
Mehrdad Yaghoobi,
In collaboration with: Sangnam Nam, Remi Gribonval, and Mike E. Davies

Workshop on Sparsity, Compressed Sensing and Applications,
Centre for Digital Music, Queen Mary University of London, UK
November 5th, 2012

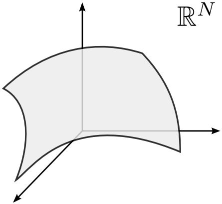
Low-dimensional Signal Models



Union of Subspaces Model



Point Cloud Model



Smooth Manifold Model

Some Applications of Low-dimensional Models

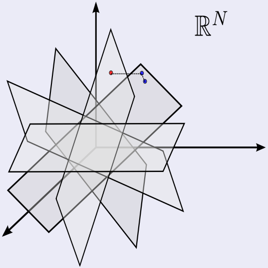
Denoising:

Contaminating with noise,

$$\mathbf{z} = \mathbf{y} + \mathbf{n}.$$

Denoising:

$$\mathbf{y}^* = \mathrm{argmin}_{\theta \in \mathcal{U}} \|\mathbf{z} - \theta\|^2$$

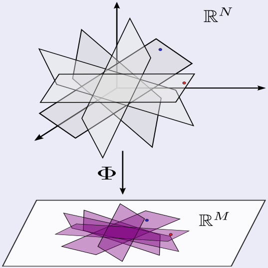


Embedding:

Embedding to a lower-dimensional space $\mathbb{R}^M$, using Φ, *i.e.* $\mathbf{z} = \Phi\mathbf{y}$.

Recovering:

$$\mathbf{y}^* = \mathrm{argmin}_{\theta \in \mathcal{U}} \|\mathbf{z} - \Phi\theta\|^2$$

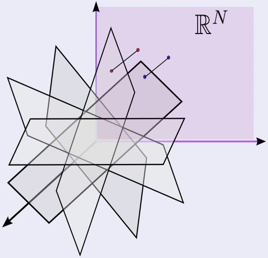


Inpainting:

Masking the signal with $\mathbf{M}$, *i.e.* $\mathbf{z} = \mathbf{M}\mathbf{y}$.

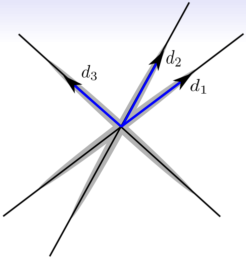
Inpainting:

$$\mathbf{y}^* = \mathrm{argmin}_{\theta \in \mathcal{U}} \|\mathbf{z} - \theta\|^2_{\mathbf{M}}$$



Sparse Synthesis Model

- Each subspace can be interpreted as the span of a small number of atoms.
- This model can be used to *represent* the signals.
- It has been used for many applications, particularly for regularising inverse problems.



Synthesis Sparsity Model

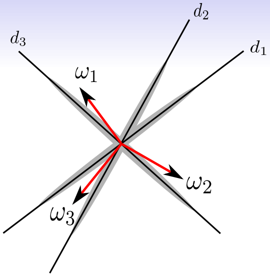
The signal $\mathbf{y}$ follows the model, if there exists an (overcomplete) dictionary $\mathbf{D} \in \mathbb{R}^{n \times p}$, $p \geq n$, such that $\mathbf{y}$ can be represented by,

$$\mathbf{y} = \mathbf{D}\mathbf{x},$$

where $\|\mathbf{x}\|_0 = k$ and k is called the sparsity of $\mathbf{y}$, in $\mathbf{D}$.

Cosparse Analysis Model

- Each subspace can be interpreted by a set of normal vectors.
- This low-dimensional model is based on constraining the possible signals.
- It has often been used for denoising.



Analysis Model

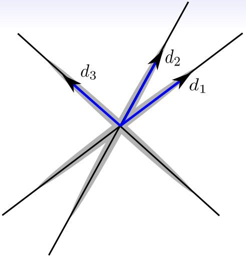
The signal $\mathbf{y}$ follows the model, if there exists a (linear) analysis operator $\Omega \in \mathbb{R}^{a \times n}$, $a \geq n$ that sparsifies $\mathbf{y}$,

$$\mathbf{z} = \Omega \mathbf{y}.$$

$\|\mathbf{z}\|_0 = a - q$, where $q > 0$ is called the co-sparsity of $\mathbf{y}$, with respect to Ω.

Dictionary Learning

- A set of exemplars $\mathbf{Y} = [\mathbf{y}_1 \ \ldots \ \mathbf{y}_i \ \ldots \ \mathbf{y}_L]$ is given.
- The goal is to find a suitable dictionary for the synthesis sparse representation of training samples.
- The dictionary is often learned by minimising an objective which simultaneously sparsify the solution and reduce the fidelity of sparse representation.



Learning Formulation

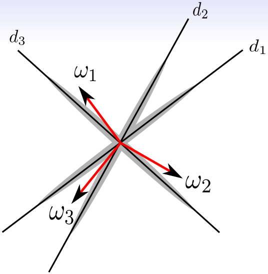
The dictionary can be learned by minimising an objective based on $\mathbf{X}$ and $\mathbf{D}$,

$$\min_{\mathbf{X},\mathbf{D}} \|\mathbf{X}\|_1 + \frac{\lambda}{2}\|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 \ \text{s.t.} \ \mathbf{D} \in \mathcal{D}$$

The constraint $\mathcal{D}$ is necessary to resolve the scale ambiguity.

Analysis Operator Learning (AOL)

- Similarly, the set $\mathbf{Y} = [\mathbf{y}_1 \ \dots \ \mathbf{y}_i \ \dots \ \mathbf{y}_L]$ is given.
- The goal is to find an analysis operator Ω such that $\|\Omega\widehat{\mathbf{Y}}\|_0$ is small, where $\widehat{\mathbf{Y}}$ is close to $\mathbf{Y}$.
- When the noiseless exemplars are available, the AOL is easier, as we do not need to find $\widehat{\mathbf{Y}}$.
- The objective is non-smooth $\Rightarrow$ not suitable for optimisation with variational techniques.



Formulation

The learned operator can be found by minimising the sparsity promoting operator,

$$\min_{\Omega, \widehat{\mathbf{Y}}} \|\Omega\widehat{\mathbf{Y}}\|_1 + \frac{\theta}{2}\|\widehat{\mathbf{Y}} - \mathbf{Y}\|_F^2 \ \text{ s. t. } \ \Omega \in \mathcal{C}$$

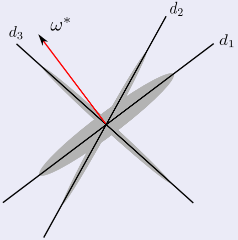
where $\mathcal{C}$ is a constraint to exclude the trivial solutions, *e.g.* $\Omega = \mathbf{0}$.

Insufficient Constraints

Row norm constraints

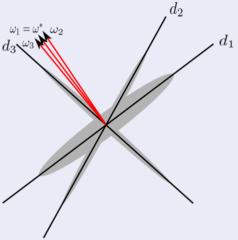
$$\forall i, \ \|\omega_i\|_2 = c$$

Rank one Ω_1 is found by repeating the best (almost) orthogonal direction ω^* to columns of $\mathbf{Y}$.



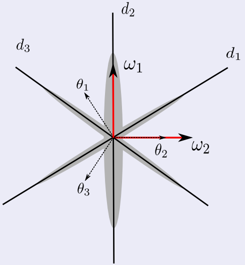
Row norm + full rank constraints

A randomly perturbed Ω from Ω_1, *i.e.* row normalised $\Omega_1 + \mathbf{N}$, has a full rank and it is still not suitable.



Tight frame constraints

It resolves the issue in a complete setting. In the overcomplete cases, it includes zero-padded orthobases.



Proposed Constraint

Uniform Normalised Tight Frame (UNTF):

Definition: $\mathcal{C} = \{\Omega \in \mathbb{R}^{n \times m} : \Omega^T \Omega = \mathbf{I} \ \& \ \forall i \ \|\omega_i\|_2 = \sqrt{\frac{m}{n}}\}$

Pros and Cons:

- Zero-padded orthobases are not UNTF.
- There exist some practical methods to project onto the TF and the UN manifolds. However, there is no analytical way to find the projection onto the UNTF!
- There is no easy way to find the global optimum, using $\mathcal{C}$ as the constraint.

Cosparse Analysis Operator Learning Algorithm

Iterative Analysis Operator Learning Algorithm

$$\min_{\boldsymbol{\Omega}, \widehat{\mathbf{Y}}} \|\boldsymbol{\Omega}\widehat{\mathbf{Y}}\|_1 + \frac{\theta}{2}\|\widehat{\mathbf{Y}} - \mathbf{Y}\|_F^2 \quad \text{s.t.} \quad \boldsymbol{\Omega} \in \mathcal{C}.$$

Solving by *alternating minimisation* technique.

- Optimisation based on $\boldsymbol{\Omega}$: Minimisation of a convex objective subject to the intersection of two manifolds $\Rightarrow$ a variant of projected subgradient algorithm is a good candidate.
- Optimisation based on $\widehat{\mathbf{Y}}$: a convex program. $\longrightarrow$ Douglas-Rachford Splitting (DRS) technique was used to efficiently solve the program.
- Here, algorithm usually converges after a few number of alternating minimisation.

Projected Subgradient Algorithm for AOL

Projected Subgradient Type Algorithm for AOL

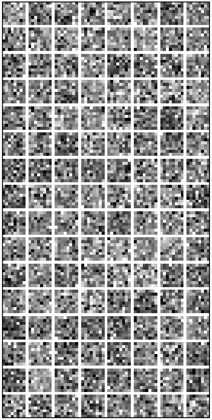
1: **initialisation:** $k = 1$, K_{max}, $\Omega^{[0]} = \mathbf{0}$, $\Omega^{[1]} = \Omega_{in}$, γ, $\epsilon \ll 1$
2: **while** $\epsilon \leq \|\Omega^{[k]} - \Omega^{[k-1]}\|_F$ and $k \leq K_{max}$ **do**
3: $\quad \Omega_G = \partial f(\Omega^{[k]})$
4: $\quad \Omega^{[k+1]} = \mathcal{P}_{UN}\left\{\mathcal{P}_{TF}\left\{\Omega^{[k]} - \gamma\Omega_G\right\}\right\}$
5: $\quad k = k + 1$
6: **end while**
7: **output:** $\Omega_{out} = \Omega^{[k-1]}$.

AOL for the Piecewise Constant Images

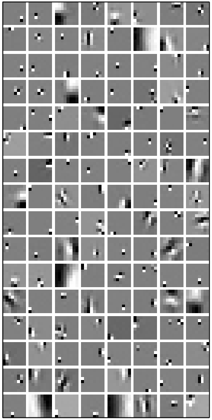
- Finding an Ω for the image patches of size 8×8.
- A 512×512 Shepp-Logan phantom image was used as the training image in a noiseless setting.
- $N = 16384$ image patches was randomly chosen from the training image.
- A pseudo-random UNTF operator $\Omega_0 \in \mathbb{R}^{128 \times 64}$ was used as the initial operator and K_{max} was 100,000.

AOL for the Piecewise Constant Images

Original Operator



Learned Operator



Issues with the Projected Subgradient Algorithm: Some Proposed Relaxations

- No analytical way to project onto UNTF $\rightarrow$ **no convergence** proof.
- Projection onto TF needs a full SVD calculation $\rightarrow$ **expensive implementation** and non-scalable algorithm.
- ℓ_1 term is not differentiable $\rightarrow$ **slow convergence** of the projected subgradient algorithm.

Relaxed AOL Formulation

1. **Relaxing the objective:** using a convex, but differentiable sparsity constraint $g(\Omega\mathbf{Y})$, where g is an entrywise function defined as,

$$g(x) = |x| - s\ln(1 + |x|/s),\ \ s \in \mathbb{R}^+, s \ll 1$$

2. **Relaxing the constraint:** using quartic constraints $\|\Omega^T\Omega - \mathbf{I}\|_F^2 \leq \epsilon_{TF}$ and $\left(\omega_i^T\omega_i - \frac{m}{n}\right)^2 \leq \epsilon_{UN}, \quad \forall i \in [1, n]$

Relaxed Analysis Operator Learning

Relaxed Analysis Operator Learing Formulation

An unconstrained objective is generate by using two Lagrange multipliers γ and λ:

$$f(\mathbf{\Omega}) = g(\mathbf{\Omega Y}) + \frac{\gamma}{4}\|\mathbf{\Omega}^T\mathbf{\Omega} - \mathbf{I}\|_F^2 + \frac{\lambda}{4}\sum_i \left(\omega_i^T\omega_i - \frac{m}{n}\right)^2 .$$

- $f(\mathbf{\Omega})$ is differentiable and it would also be convex, if we restrict its domain to $\mathcal{C}_c = \{\mathbf{\Omega} : \mathbf{\Omega}^T\mathbf{\Omega} - \mathbf{I} \succeq \mathbf{0},\ \forall i,\ (\omega_i^T\omega_i - \frac{m}{n}) \geq 0\}$.

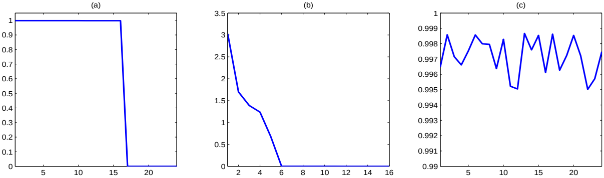
Gradient Descent Algorithm for AOL

A variable step-size gradient descent, with line search, can be used to minimise $f(\mathbf{\Omega})$, where the gradient of f can easily be found by:

$$\nabla f = \left[\frac{\mathbf{Z}_{i,j}}{s + |\mathbf{Z}_{i,j}|}\right]_{i,j}\mathbf{Y}^T + \gamma\Big(\mathbf{\Omega\Omega}^T - \mathbf{I}\Big)\mathbf{\Omega} + \lambda\left[\omega_i\left(\omega_i^T\omega_i - \frac{m}{n}\right)\right]_i^T$$

$$\mathbf{Z} := \mathbf{\Omega Y}$$

Relaxation of the Constraints



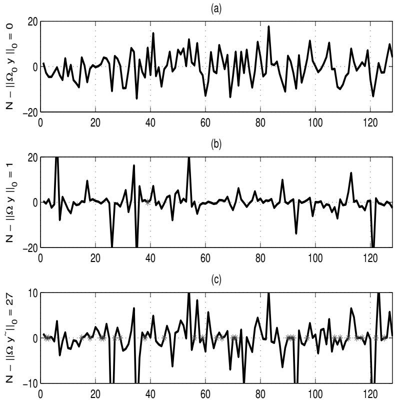
- Learning an $\Omega \in \mathbb{R}^{24\times 16}$ from $\mathbf{Y} \in \mathbb{R}^{16\times 576}$ $q = 10$ cosparse exemplars.
- Sorted ℓ_2 norms of the rows of learned operator with the TF constraint (left).
- Singular values of the learned operator with the UN constraint (middle).
- Normalised inner-products between the rows of the synthetic ideal operator and the corresponding rows in the learned operator (right).

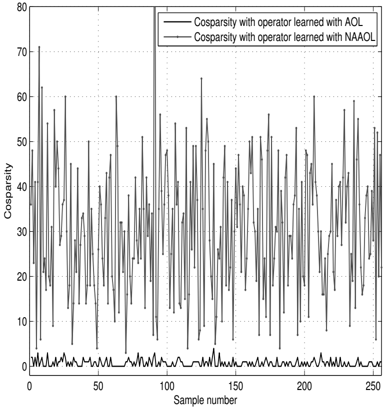
An Operator for the Face Images: Setting

- Learning an Ω for the image face patches from the Yale face database.
- $L = 16384$, 8×8 image patches were randomly selected from different faces.
- The noise-aware and noiseless AOL methods were used for operator learning.

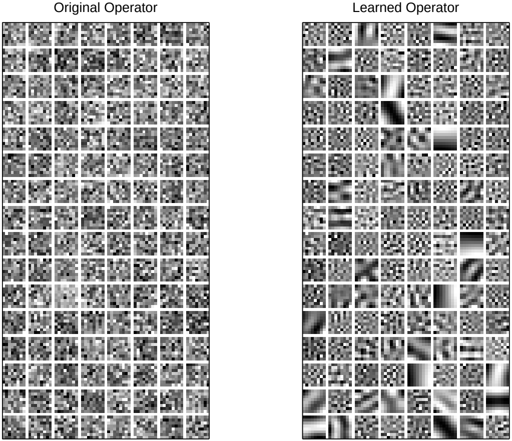
An Operator for Face Images: Cosparsity Comparison

- The analysis coefficients $\mathbf{z} = \mathbf{\Omega y}$ and cosparsities were calculated, using $\mathbf{\Omega}_0$, $\mathbf{\Omega}_{AOL}$ and $\mathbf{\Omega}_{NAAOL}$.



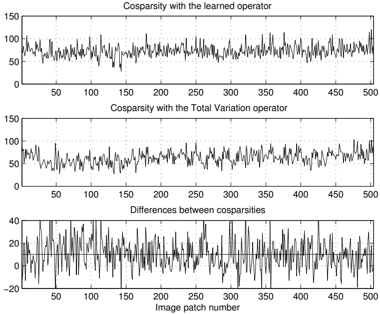


Learned Operator



Face Images Denoising: TV v.s. Learned Operator

- TV operator for comparison.
- Two different regularisation parameters, $\lambda = 0.3$ & 0.1 .



Conclusion and Future Work

Conclusion:

- The constrained analysis operator learning is a useful technique to find a suitable analysis operator.
- The proposed constraint can be relaxed to reduce the complexity of the optimisation algorithm, while including some **approximately UNTF** operators.
- The simulation results emphasis on the fact that we should use the correct analysis operator, i.e. TV or oscillatory operators.
- The convergence of the relaxed AOL is guaranteed, as its objective has a bounded curvature and its sublevel set is compact.

Future Work:

- Investigating the **local identifiability** of operators in this framework.
- More investigations on the **structures** of the learned operators.

Thanks for your attention.